The power of words: Towards a methodology for progress monitoring in design thinking projects

Dolata, Mateusz ; Übernickel, Falk ; Schwabe, Gerhard

Abstract: The popularity of design thinking as an innovation paradigm grows continuously. More and more schools and firms implement innovation processes inspired by design thinking, but they lack easy and nonintrusive methods for monitoring the progress of teams following those processes. Consequently, interventions from coaches, teachers, or supervisors tend to rely on intuition or require intensive and intrusive examination of team dynamics. This study uses by-products from the design process and proposes automated assessment of lexical diversity as a monitoring method in process-driven design thinking projects. Thereby, it contributes to the research on the relation between text production and creativity in design projects. To the practical end, it suggests how digitalized by-products of design activities such as notes and documentation, can be leveraged to support the teams as well as coaches, teachers, and supervisors.

Originally published at:
Dolata, Mateusz; Übernickel, Falk; Schwabe, Gerhard (2017). The power of words: Towards a methodology for progress monitoring in design thinking projects. In: 13th International Conference on Wirtschaftsinformatik 2017, St. Gallen, Switzerland, 12 February 2017 - 15 February 2017, AIS Electronic Library.

The power of words: Towards a methodology for progress monitoring in design thinking projects

Mateusz Dolata[1], Falk Uebernickel[2], and Gerhard Schwabe[1]

[1] University of Zurich, Department of Informatics, Zurich, Switzerland
{dolata, schwabe}@ifi.uzh.ch
[2] University of St. Gallen, Institute of Information Management, St. Gallen, Switzerland
falk.uebernickel@unisg.ch

Abstract. The popularity of design thinking as an innovation paradigm grows continuously. More and more schools and firms implement innovation processes inspired by design thinking, but they lack easy and nonintrusive methods for monitoring the progress of teams following those processes. Consequently, interventions from coaches, teachers, or supervisors tend to rely on intuition or require intensive and intrusive examination of team dynamics. This study uses by-products from the design process and proposes automated assessment of lexical diversity as a monitoring method in process-driven design thinking projects. Thereby, it contributes to the research on the relation between text production and creativity in design projects. To the practical end, it suggests how digitalized by-products of design activities such as notes and documentation, can be leveraged to support the teams as well as coaches, teachers, and supervisors.

Keywords: design thinking, lexical diversity, documentation, creativity in teams, progress monitoring

1 Introduction

Establishing innovation culture in education and industry comes along with propagation of specific problem-solving strategies including particular tools and processes showing when and how to apply those tools. This holds also for a continuously booming innovation paradigm: design thinking (DT) – firms and schools around the world borrow from each other and propagate “successful” DT process models and implement them in their own innovation departments and classes [1–3]. They struggle, however, with an ongoing monitoring of the design teams’ progress along the process [4]: Do the teams go the right path? Do they make enough improvement from one phase to another? When to intervene and when not? This paper shows the power of the words used in the documents produced by the teams: using metrics based on the vocabulary diversity bears great potential for identifying design team’s progress without expensive external assessment or self-reflection episodes.

A constantly growing community of teachers, practitioners, and researchers applies DT and similar, process-dependent methods when solving wicked problems. They benefit from the agility and flexibility of the methods, as well as from the fragmentation of

Manuscript accepted for presentation at Wirtschaftsinformatik 2017

the problem and solution space enforced by the particular tools and processes. Consequently, the problems and possible solutions become more tangible, comprehensible, and user-specific. Despite large number of handbooks and guides for DT [1, 5, 6], the practitioner-oriented community still lacks deep understanding of what actually happens during a DT project. This knowledge would support the coaches and supervisors at monitoring the progress of design teams as they follow the process and at proposing meaningful interventions. As for now, they mostly rely on intuition and experience – based on the findings, we claim that their tasks could be even partially automated. Additionally, generalized set of observations regarding DT would also help the researchers with creating abstract toolset-independent models of design and creativity, which could be then transferred to novel contexts. Since the community around DT is growing continuously, we expect that the discussion on the related topics will intensify [7].

As a paradigm, DT relies on the design, production, and analysis of artifacts [1, 2]. During workshops, temporary notes capture ideas in form of easy drawings and simple descriptions for prototyping. Prototypes capture ideas in a tangible, plastic form for testing. Observations and intermediate results collected in tests are recorded as voice or text for later analysis. And, finally, results of the analysis are made persistent as text or diagrams for use in design workshops. Thanks to the recent technological progress, most of the data is shared virtually across team members, gets regularly archived in common repositories, and has textual character or format which can be automatically turned into text (e.g., voice recordings via auto-transcription or photos via character recognition). Consequently, we find ourselves in a situation where, on the one hand, organizations struggle with monitoring design teams, and on the other hand, those teams produce and file more and more potentially relevant and easy-to-process textual data. As explained below, previous research found a link between lexical diversity and the creativity of a narrative – in an educational context not related to any project group work or design documentation [8]. It also provides some evidence, that textual analysis of final documentation produced by design teams helps assessing their overall performance [9, 10]. However, as the literature leaves open whether and how those insights can be used for monitoring purposes in long-term projects, we hypothesize that *vocabulary diversity of documents produced within a DT project exhibits patterns which reflect the DT process* and, consequently, can be used to monitor the progress of design teams with regard to this process.

This study handles the above hypothesis, while analyzing a set of data produced by 15 design teams in an educational context in 6 continuous years. First, the paper discusses the process-based view on DT and ways proposed to monitor progress in similar projects; it also motivates usage of language metrics as a monitoring instrument. Second, it discusses the methodology used including the metrics used for the current analysis. Next, it provides an overview of results followed by the discussion thereof in light of what we know about DT and progress measurements. Finally, it summarizes the insights and explicates limitations of the study along with directions for future work.

2 Related Work

2.1 Design Thinking as a process

Design thinking (DT) is an innovation paradigm that gains increasing popularity and reputation. It originates from the academic world, where it was conceptualized as a teaching methodology in the area of mechanical engineering [2, 7, 11]. Later on it found its way to industry and is applied to solve practical problems in a variety of industries [5, 7, 12]. Successes of consulting firms like IDEO popularized DT as a mindset – institutions who wanted to replicate its success required, however, also a more hands-on guidance [13]. Consequently, DT was formalized as a toolset including a family of process models to be applied in previously unknown contexts [1–3, 11, 14].

Available literature proposes various formulations of the DT process. Large part of the models focus on the iteration through divergent and convergent phases when developing multiple prototypes [2, 5, 13]. More advanced models introduce blocks to differentiate between user-centered (e.g., empathize, define) and prototype-centered (e.g., ideate, prototype, test) [6]. Yet other models propose a chronological process built around phases that formulate specific requirements at the prototypes developed therein [1, 3, 11, 15]. The availability of phase-specific requirements and goals suggests that it is possible to monitor teams' progress through comparing prototypes produced in each phase with the phase goals. However, this kind of assessment requires extensive human power, especially in cases where large numbers of prototypes are produced. Also, this method is prone to subjectivity. This leads to the question whether a more objective and less extensive way of monitoring is possible.

This study focuses on the DT process model established in mechanical engineering education at Stanford [11] and adapted in a multi-national community of universities [1, 3, 16]. The macro process consists of seven milestones divided in two large phases: the divergent phase and the convergent phase. The milestones in the divergent phase include [1]: (1) *design space exploration* (DSE) – understanding the problem space and topics related to the creative challenge, (2) *critical function prototype* (CFP) – defining the principal needs of the stakeholders and addressing them through prototypes, (3) *dark horse prototype* (DHP) – reframing the challenge and questioning previously obtained insights through further "crazy" prototypes. The divergent phase closes with (4) *funky prototype* (FKY), where most promising results are compiled into complex and open prototypes. The funky milestone forms a bridge to the convergent phase consisting of the following milestones: (5) *functional prototype* (FNL) – providing a preview for the final prototype and its key aspects, (6) *X-is finished prototype* (XFP) – designing and developing a single function of the final prototype to the last extent, (7) *final prototype* (FIN) – integrating successful elements from previous milestones into single, coherent, and high-resolution prototype. Throughout the divergent phase, teams develop large numbers of low-resolution prototypes, while in the convergent phase they invest time into specifying the scope and functionality of a single prototype. Along this macro process, teams engage in multiple design iterations (micro-process) consisting of such steps as [1]: (1) need-finding and instant expertise, (2) brainstorming, (3) prototyping, (4) testing, and (5) re-defining the problem. Steps (1) and (2) have clearly a divergent character, steps (3), (4), and (5) prompt the teams to reflect and converge.

Adherence to the process is one of the essential success factors in DT projects [17]. Consequently, monitoring the progress of teams in the macro process seems to be essential for the team itself, as well as their facilitators, managers, partners, and sponsors. Even though the process provides specific prototype requirements that could be used for a kind of assessment-based progress monitoring, it would require additional human power and could lead to subjective assessment. Publications which propose particular process models leave the question open, how to monitor their implementation in an efficient and effective way [1, 5, 6].

2.2 Monitoring design process

In most general sense, monitoring involves being aware of the state of a system. With regard to DT, by monitoring we mean whether the DT-typical mindset and toolset [14] is understood and properly followed by the team. This monitoring may address interpersonal, relational, performance, or process level. The role of monitoring in DT is essential: First, it shall enable ongoing feedback regarding the facilitation success [18]. Second, it shall help with assessment of the team's performance [19]. Consequently, being aware of the team's state and its performance with regard to the process model has an essential character for a process-dependent DT implementation.

A simple measurement of creative performance may involve characterization and statistics regarding the generated ideas [20–23]. This monitoring method uses such concepts as complexity of ideas, their amount, originality, etc. While well applicable to controlled group assignments in a test setting, the adaptability of such measurements to DT seems questionable. Complex prototypes often involve several ideas, which may be new or a specification of previous ones [24]. Consequently, monitoring using statistics about ideas did not find much attention in the DT community.

Another thread of research focuses on the assessment of the social dynamics in design teams and social characteristics thereof. Two approaches dominate in this thread: psychometrics and affect-oriented measurements. Psychometric approaches rely on such tools as the Myers-Briggs Type Indicator (MBTI) [25] or similar self-reflection tools. They have been successfully employed for team composition and prediction of its outcome [26–31]. Observations of team dynamics without a self-reflection loop rely on the availability of a trained, external observer. Their potential has been confirmed

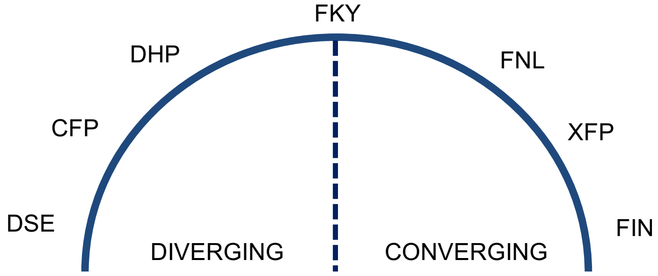

Figure 1. Phases and milestones in a DT project – chronologically from left to right [15].

with a set of studies especially from the DT community [32–35]. While team dynamics and psychometric compatibility has impact on the teams outcome, it does not help with monitoring team's progress in a process. In fact, its results may be misleading: a team that established excellent team dynamics for an early phase in the process (e.g., CFP) and does not make any progress beyond this phase, will be identified as successful by the above methods. Additionally, self-reflection and independent observation are expensive – they require dedicated time from the team or the observer. They may also be perceived as intrusive by the team. While being a well-promising idea in a laboratory, psychometric and team dynamics monitoring leaves open multiple questions on its applicability in real-world projects.

Yet another way to go involves analysis and understanding of all kinds of artifacts used or produced within creative teamwork [36–38]. This fits into the paradigm of multimodal analysis of human-generated content and actions [39]. While this approach catches the complexity of processes in artifact-oriented creative activities such as design, it has not yet turn into practical applications, thus remaining at the abstract level of frameworks and models.

To conclude, the growing popularity of DT leads to increasing attention in research. The various process models and their role for the success of DT have been extensively addressed. In parallel, several options for monitoring the team's progress with regard to the process have been proposed. However, a non-intrusive monitoring technique is still missing. Nevertheless, some preliminary studies relate language use of design teams and their overall performance [9, 10, 40, 41] – we take them as an inspiration and discuss their potential for process *monitoring*.

2.3 Language in creative processes

Language does not only describe the reality, but also "makes" it (e.g. in form of performative utterances [42]). Three primary roles of language in design can be defined: representational (explaining design process and design rationale), instrumental (communicating and recording design intent and rationale), and constructive (contributing to the design process) one [40]. Spoken language is used to describe things (ideas, concepts, information chunks) that enter the design process, while written language is used to describe things that should stay in the process or that could be useful at a later stage [10, 40]. This opens the possibility to analyze written language as a way to assess and monitor the outcome teams produce.

Language, beyond supporting and constituting design activities, is an essential element of creative processes, i.e., inventions. Vivas [43] writes that what creative persons do is "not to invent something new, but to extricate out of the subject matter at hand its own proper structure or order" – this structure is given by language in communication [8]. Several studies empirically confirm the structural nature of creativeness [44, 45]: the structure such as language drive and inform creative performance. Consequently, we argue that *language* and *creative performance* mutually influence each other – patterns discovered in one shall be reflected by specific patterns in the other one: if the creative performance is driven by the process, the language-features will be likely to reflect this process too. However, the studies that inform our argument leave it open, which characteristics of language shall be considered.

The link between creative performance and narrative performance is supported by the discourse in the area of language acquisition psychology. Narrative performance and creativity both share a number of qualities, particularly if it comes to divergent thinking [46]. Creativity, and, particularly, divergent thinking require, among others, ability to produce wide variety of ideas (flexibility) and ability to produce unusual ideas (originality). These two qualities are, also, proposed as sources for wide range of vocabulary in narrative tasks [46–48]. Lynch and Kaufman [8] identify three main factors (out of 7 main and 24 sub-factors) that have moderate factor loadings on creativity assessment: verbal diversity, readability, and consistency, whereas the first one has the highest factor loading. In other words, human judges as generally more creative perceive texts that exhibit higher diversity. Importantly, in the above studies text composition / text judgment was the main task given to the test persons. However, DT is not about writing texts, but about thriving innovative ideas. Documentation is mostly considered only a side-product of teams' activities and communication is not a goal by itself, but should support ideation and creation of the intermediate and final prototypes. Therefore, it remains open whether regularities regarding lexical diversity will hold for text others than narratives intended to be creative.

Accordingly, language of collaboration and its relation to team's creative performance has been approached within the DT community. Already in 1996, Mabogunje and Leifer [9] show that design teams in the area of mechanical engineering who introduce novel noun phrases in the mid-term and final documentation tend to produce better and more creative results than teams less innovative in their language use. Dong [10, 40] who, also, focuses on design teams in the area of mechanical engineering shows that other qualities of written language, such as coherence and presence of sentimental statements also correlate with teams' performance. However, studies leave it open how to monitor the progress of teams along a predefined process and how to use the metrics for continuous assessment rather than post-mortem analysis of projects and teams.

Overall, the role of language and evolution of language is highly relevant for creative performance and collaboration; however, it remains underspecified how language evolves over time in design teams and how it thus can be used to guide facilitation efforts. Moreover, longitudinal studies of creative performance in teams are rare, but very valuable and needed [49]. Already in 1984, McGrath [50] appealed for closer consideration and deeper research on longitudinal team collaboration. His call still remains up-to-date, in particular if it comes to teams working on creative tasks in organizational or nearly-organizational context [51–53].

3 Methodology

The context of our study is a graduate course at a Swiss University. In this class, students are taught to apply design thinking on a real-life innovation challenges, offered by industry partners. Challenges originate from different market sectors and reflect the real needs of the involved industry partners – each team works on a different challenge.

The design teams consist of graduate students with multidisciplinary backgrounds, as well as prior working experience. Each team receives coaching from DT experts. Additionally, industry partners nominate an employee as a company-internal contact partner for the team. During the course of 10 months, the teams pass the milestones described above in order to finish their projects and provide a final prototype.

The teams experience co-located and distributed collaboration. Except for the sessions with the coaches and facilitators twice a week for ca. 2 hours, the teams are free to organize their collaboration process. Sometimes, the routines get adjusted throughout the process – teams tend to intensify synchronous and co-located collaboration towards the end of the process and shortly before the particular milestones. Teams are free to choose media and technologies to support their collaboration effort. A common and non-interchangeable element for all teams forms a media wiki used as knowledge base. It was introduced into the course to enable for cross-inspiration and cross-pollution between teams within a year's class and, also, between different years' classes.

Each team regularly updates their wiki section. It includes basic information on the team, the industry partner, and the challenge. Furthermore, teams document their progress in wiki while describing (1) insights obtained through ethnographic user studies, (2) ideas implemented as prototypes along with evaluation results, and (3) experiences collected through collaboration with other teams or in research trips. They upload photos, videos, and other material into their shared space. Established in 2008, the wiki has become a large repository of design knowledge from several generations of students in this course. As of 2016 the wiki includes information for 38 challenges in German or English. For this study, we consider a subset of 15 challenges from years 2008-2015 documented in German. The overall number of documents considered equals to 630.

We decide to use the German part of the wiki because large majority of the course participants have been German or Swiss German native speakers (more than 95%). Consequently, those students who decide to run their wiki section in English, actually, choose to use a foreign language. Aware of the differences regarding lexical diversity and other related characteristics between the mother tongue and second or third language, we deliberately exclude documents written in English from this study.

To assess the lexical diversity, we employ MTLD (Measure of Textual Lexical Diversity) [54]. While, in general lexical diversity refers to ratio of total number of words to the number of different unique word stems, we decide to use the specific metric of MTLD for a particular reason – according to the empirical tests run by McCarthy and Jarvis [54] it is resistant to length-induced variance in texts of 100 to 2000 words. 98% of documents considered in this study are of this length. As most measurements for lexical diversity, MTLD takes Types-Token Ratio (TTR) into consideration, however it expresses the mean length of sequential word strings with a given TTR value (the optimal value of this factor was empirically estimated to equal to 0.7 [54]). In other words, it provides information on the average text length (measured by number of words) with a TTR above the used treshold – if the results is 100, it means that on average after each 100th word the TTS value dropped below the threshold of 0.7. To assure the measurement to be robust, the MTLD algorithm applied in this study does a forward and backward run through the text to compute the final outcome [55].

4 Results

Comparison of lexical diversity of documents from the divergent phase (milestones: DSE, CFP, DHP) and the convergent phase (FNL, XFP, FIN) shows a clear picture: while mean lexical diversity in the divergent phase equals to 242 (Standard Deviation SD = 67, Standard Error Mean SE = 3), in the convergent phase MTLD yields on average an outcome of 192 (SD = 78, SE = 10). The difference is significant and accounts for $p < 0.001$ ($t = 4.73$, $df = 73$; in a two-tailed unpaired Student's t-test without assuming equal variances – this test was chosen due to unequal samples – see below).

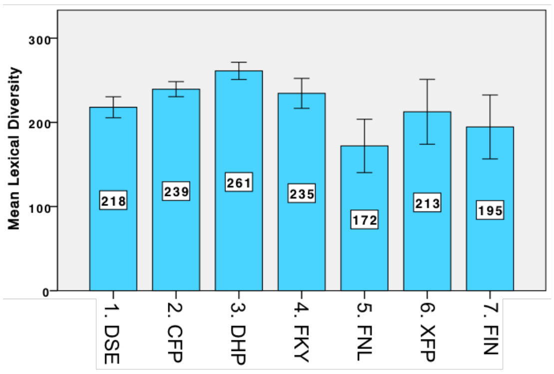

Figure 2. Mean lexical diversity (measured with MTLD) of documents for each milestone in the DT process. Data considered for 15 projects (630 documents). Error bars: 95% CI

Comparison of lexical diversity for subsequent milestones allows for analysis of process-typical tendencies. For easier comprehension, Figure 2 includes the chart representing the mean values along with error bars. In the divergent phase, we can identify a growing tendency – lexical diversity increases significantly from milestone to milestone. MTLD algorithm yields the following values: for documents in DSE it yields 218 (SD = 66, SE = 6), in CFP – 239 (SD = 63, SE = 5), and in DHP – 261 (SD = 69, SE = 5). For each step, we compute a Student's t-test and provide its results in Table 1.

FKY is the first milestone with a drop in the lexical diversity: mean lexical diversity of documents equals to 235 (SD = 81, SE = 9). However, the largest drop occurs afterwards, i.e., when the mode changes to converging. MLTD computation for the milestones in the convergent phase average to the following values: FNL – 172 (SD = 71, SE = 15), XFP – 213 (SD = 84, SE = 18), FIN – 195 (SD = 19, SE = 18). As presented in Table 1, no significant difference between those milestones could be identified – while this may be an effect of smaller number of documents considered for the statistical analysis (cf. Table 1; teams produce fewer documents as they converge and exhibit tendency to collect information in few central documents), we can definitely see that none of the phases in the convergent phase exhibit higher lexical diversity than any of the phases in divergent mode. Additionally, no clear tendency – neither growing nor lessening – can be identified for the convergent phase. We attribute the difference regarding the standard error (reflected by the length of error bars) to the number of documents teams produce in particular phases: in the divergent phase, the number of distinct documents is higher, while in the convergent phase the number of distinct documents is lower; consequently, the average figures on document lengths for each phase in each team lie further away from each other in the divergent phase.

Milestones and their short forms

Design Space Exploration
DSE (111)

Critical Function Prototype
CFP (193)

Dark Horse Prototype
DHP (174)

Funky Prototype
FKY (83)

Functional Prototype
FNL (22)

X-is Finished Prototype
XFP (21)

Final Prototype
FIN (19)

Considered milestones	**Results of a T-test regarding comparison of mean lexical diversity for the two considered milestones**		
	t	**df**	**p**
DSE → CFP	2.744	220	.006 *
CFP → DHP	3.154	353	.002 *
DHP → FKY	2.579	139	.011 *
FKY → FNL	- 3.53	36	.001 *
FNL → XFP	1.691	39	.100
XFP → FIN	- 0.69	38	.491

Table 1. Results of the t-test for lexical diversity in subsequent milestones. On the left: list of all milestones with their respective short forms and number of documents considered.

To confirm the significance of the results that can be clearly seen in Figure 2 (by comparing the error bars), we decided to run a Student's t-test. Due to different number of documents in each phase in each team and because we did not want to exclude any documents from the analysis, we decided to employ a two-tailed unpaired t-test without assuming equal variances – the results for each subsequent pair of milestones can be seen in Table 1. We also run a two-tailed paired t-test where we used samples of 19 random documents from each milestone to assure that measured effects do not occur simply due to the large number of documents or difference between milestones in this regard. The results of those t-tests confirmed the significance of the results while yielding p-values $p<0.05$ in large majority of cases for milestone pairs marked with an asterisk in Table 1, i.e., those where difference regarding lexical diversity is significant.

5 Discussion

The results presented above confirm the hypothesis expressed in the introduction: lexical diversity of texts produced in a DT project can deal as a metric to monitor the progress of a team with regard to the DT process. The following reiterates this finding while pointing to the consequences it has for the monitoring of DT activities.

5.1 Particularities of the DT process

We observe, that the lexical diversity of the documents grows significantly between the milestones in the divergent phase – and reaches its peak in the dark horse phase, where teams are encouraged to leave behind previous ideas and consider totally new ideas and approaches [1, 2]. Dark horse prototypes enjoy a special status in the design thinking community – they are often presented as the source of final ideas and solutions [17]. With the above results we confirm that this milestone is, in fact, special – also with regard to the lexical diversity: it is clearly the milestone that exhibits the significantly highest lexical diversity out of all milestones. Interestingly, as soon as first consolidation efforts are made – in the funky milestone, where teams combine knowledge from previous milestones without the necessity to focus on a single coherent prototype – a slight drop in the lexical diversity occurs. We claim, that this reflects the fact, that they drop some ideas and concepts and tend to combine the most promising ones.

However, the most dramatic drop in lexical diversity occurs, where teams are pushed to converge while focusing on a single prototype built around coherent concepts. When considering the role of written language in DT process as a medium to retain things for later reference (representational and instrumental-recording role [40]), the changes in lexical diversity reflect the evolution of solution space. While in the divergent phase, the scope of potential solutions is rather large and should grow even further, in the convergent phase it is clearly limited. It is the time, where teams change their work modus from creativity to execution [1, 11] and they assess their own performance rather in terms of completing predefined tasks than in providing unexpected solutions [11]. The lack of significant differences during the convergent phase has possibly two reasons: (1) during that time, teams generate fewer documents as they concentrate on building their prototypes and do not produce new knowledge that needs saving; (2) the milestones are not as distinct as in the divergent phase – they merge into one another. Nevertheless, also in the convergent phase some variance occurs. While not being significant, the difference between FNL and XFP phases points to an interesting aspect of the DT process: the functional prototype focuses on a set of key concepts and is mostly built around a story which allows for preview of the final functionality – a successful story tends to be coherent and concise, however in XFP teams have small divergent elements, e.g., when they look for and decide between possible technical solutions to implement the single, “X” feature. Consequently, the changes of lexical diversity between particular phases reflect how goals and tasks evolve throughout a DT project.

Importantly, the projects we consider for the current study, follow a very popular DT process. Through its rigidity and the available hands-on guidance [1], it enables for a relatively easy transfer from one context to another. Consequently, it attracts attention from educational sector as well as from the industry [11]. We claim, that the presented results contribute even further to the better understanding of this particular process and its building blocks or phases. Though, it remains open, whether similar patterns can be identified in projects following another DT process model [2, 5, 6, 15].

5.2 Lexical diversity in creativity

Our findings confirm the interrelation between language production and creativity. In such complex processes as group creativity it is almost impossible to make conclusive hypothesis regarding causal relation between variety of words and divergent thinking.

Does divergent thinking cause the lexicon growth, or does lexicon growth cause divergent thinking? In accordance with previous studies [8, 42–45], we argue that language has a structural nature and, along with the predefined process, it gives a specific structure to the creative efforts of a design team [42, 43]. With this in mind, providing means to enrich the vocabulary of design teams may positively influence their creative performance in divergent phase, while reducing stimulation may positively influence team's performance in convergent phase. This opens possibilities for design of stimulation systems with specific focus on the vocabulary.

The currents study provides insights that go beyond the previous literature in twofold direction. First, it provides evidence for the link between the creative effort and textual production, even if production of the text does not constitute the central task during this effort. Consequently, it extends the findings presented by Lynch and Kaufman [8]. While our study points to the link between lexical diversity and creativity in a longitudinal project context, it leads to the question whether similar effects can be observed in local context: Do descriptions of "more creative" prototypes tend to use more diverse vocabulary? Answering this question suggest a direction for further research.

Second contribution this study makes to previous research on relation between lexicon and creativity relates more to the DT context. Mabogunje and Leifer [9] as well as Dong [10, 40] provide evidence for the relation between the overall assessment of the project (done after the project) and specific linguistics qualities of documentation (sentimentality, coherence, appearance of neologisms). We extend the list of relevant characteristics by pointing to lexical diversity as a possible and easy-to-apply metric. As opposite to previous research [9], which focused on arbitrary division of time (limited to three academic quarters) and related the results primarily with academic grades, our solutions provides higher-granularity observations on the teams' progress across the DT process. Additionally, the method proposed in here does not require extensive preprocessing [9] and can be applied to documents with a short length, which, normally, include only few noun phrases. Application of the proposed metric for performance monitoring would require further research, as discussed above.

5.3 Lexical diversity and monitoring of creative processes

Resilient methods for monitoring design processes are now necessary more than even before, as the popularity of agile design and innovation methods grows particularly fast. Using texts for assessing design teams progress provides a way to monitor DT teams without steady human control and without use of such intrusive tools as surveys and other psychometrics. In line with this, teams tend to produce and digitalize more and more data, especially side products from their design activities. This opens options going beyond the ones suggested in previous literature.

Monitoring and interventions relying on observing social dynamics of teams is extensively propagated within the DT community and is considered one of the central tasks of teams coaches [11, 17, 32–35]. Similarly, much attention is put on the controlled, conscious team composition and its effects [26–31]. Those efforts require very intensive observations, good intuition, and readiness to intervene from the coaches, thus compromising teams' autonomy and distracting them from the actual task. Even though we accept arguments for intensive coaching of innovation projects, especially with unexperienced designers, with lexical diversity we point to a source of additional, independent data useful for coaching. Since such measurement can be done independent of the team and does not require extensive human power from the coach, nor his or her intuition, it offers an objective and nonintrusive way of monitoring.

Monitoring based on statistics regarding number of singular ideas [20–23] or various qualities of prototypes [36–38] provides another way for nonintrusive process monitoring. While well applicable to particular brainstorming or prototype-design sessions, those methods are hardly applicable for DT, because prototypes and artifacts used therein are often complex and unusual. Comparing them directly with each other, even within a single project, may feel like comparing the incomparable. Under those circumstances, focusing on and monitoring side products of the design process seems reasonable. While this study provides evidence for a relevant role of vocabulary and its diversity, we claim that more metrics can be applied in similar way, just to mention argument structure in documents, cohesiveness, sentimentality, or comprehensibility.

Automatizing the method described in this study are possible directions for future design efforts to establish, e.g., a DT coach cockpit. Such a cockpit could provide on-the-go analyzes of documentation generated by the team, as well as their communication data (most teams use, e.g., WhatsApp to discuss or document ideas). The central insight from the study to be applied in such a system are the lexical diversity differences between phases. For instance, if the system detects a drop of lexical diversity in the textual material produced by a particular team between the CFP and DHP, i.e., a signal of converging tendencies in a divergent phase, the coach may intervene by providing more diverse stimulation material, introducing additional assignments, or organizing brainstorming workshops. The system may support the coach by monitoring other relevant information, such as the structure of documents or the relation between them [43, 45, 46]. Another point to observe would be the number of meetings between team members to assess whether possibilities for ideas transfer occurred at al. As for now, coaches intervene based on their subjective observation intuition, and are limited to the face-to-face meetings with the team. An independent metric, lexical diversity, makes it possible to change this punctual monitoring routine to a continuous monitoring, which in turn allows for more precise and swift interventions as well as more dedicated coaching.

6 Conclusion and Limitations

To conclude, this paper confirms the intuition inspired by earlier research in design engineering [9, 40] and in text creativity [8, 46, 47] that evolution of vocabulary in design documentation reflects the process models applied in such collaboration. Following this finding, one can speculate about the usage of documents produced by the team as a source of knowledge on team's progress. It also confirms the strong connection between the language we use and our state-of-mind. If a mutual relation between the two is assumed [43], then stimulating language will enhance creative performance, and experiencing creativity will result in observable changes in language. Consequently, both dimensions should be considered in collaborative systems designed to support creative and divergent thinking. Our findings as well as the discussion shall inspire the scientific community to reconsider the topic of creativity support, in particular, with regard to long-term innovation projects. The DT community benefits from better understanding of the process model it promotes and how it influences the communication in design teams. Finally, facilitators of design projects may use the findings to control the progress made by teams, both in organizational and educational context.

While we postulate that language can be used for monitoring and stimulation of creativity in design teams, we acknowledge the fact, that language, and particularly, lexicon is not the only important factor in this regard. We would see it as ethically doubtful, if, e.g., university grades were depending on the lexical diversity or similar measurements. In fact, linguistic performance of an individual may vary and depends on daily mood, psychophysical state, and other independent factors, such as team composition.

Additionally, this study concentrates on the differences between phases in a "normal" case, simulated by the average measurements of all teams – comparison of lexical patterns between very successful and less successful project would provide further support for monitoring. Beyond that, it would allow for performance assessment of teams and not only their adherence to the process. To conduct such study, we would, however, require extended material – dividing 15 teams in two or more groups depending on their grades results in small "n" and yields results with only limited statistical significance. We therefore call for further research in this area with a more extended data set. Also, establishing the relation between the lexical diversity and the actual creativity of the ideas described in the considered documents would provide additional evidence.

Another related question addresses the influence of mother tongue and second/third language on the linguistic performance in design process documentation. We concentrated on the German texts in this study and disclosed wiki sections written in English, i.e., in a foreign language for the authors and presumed audience. However, it would be definitely an important extension of the current study to run the same analysis on data generated by native speakers of other languages.

References

1. Uebernickel F, Brenner W, Naef T, et al (2015) Design Thinking: Das Handbuch. Frankfurter Allgemeine Buch, Frankfurt am Main, Germany.
2. Skogstad P, Leifer L (2011) A Unified Innovation Process Model for Engineering Designers and Managers. In: Meinel C, Leifer L, Plattner H (eds) Des. Think. Springer, pp 19–43
3. Uebernickel F, Brenner W (2016) Design Thinking. In: Hoffmann CP, Lennerts S, Schmitz C, et al (eds) Bus. Innov. St Galler Modell. Springer Fachmedien Wiesbaden, pp 243–265
4. Bushnell T, Steber S, Matta A, et al (2013) Using A "Dark Horse"Prototype to Manage Innovative Teams. 3rd Int. Conf. Integr. Des. Eng. Manag. Innov.
5. Brown T (2008) Design Thinking. Harv Bus Rev 86:84–92.
6. d.School (2015) Design Thinking: The d.School Bootcamp Bootleg.
7. Johansson-Sköldberg U, Woodilla J, Çetinkaya M (2013) Design Thinking: Past, Present and Possible Futures. Creat Innov Manag 22:121–146.
8. Lynch MD, Kaufman M (1974) Creativeness: Its Meaning and Measurement. J Lit Res 6:375–394.

9. Mabogunje A, Leifer LJ (1996) 210-NP: measuring the mechanical engineering design process. In: Front. Educ. Conf. 1996 FIE96 26th Annu. Conf. Proc. Of. IEEE, pp 1322–1328
10. Dong A, Hill AW, Agogino AM (2004) A Document Analysis Method for Characterizing Design Team Performance. J Mech Des 126:378.
11. Carleton T, Leifer L (2009) Stanford's ME310 Course as an Evolution of Engineering Design. Proc CIRP Des. Conf.
12. Brenner W, Witte C (2011) Business Innovation CIOs im Wettbewerb der Ideen ; mit der Methode Design Thinking. Frankfurter Allg. Buch, Frankfurt am Main
13. Brown T (2009) Change by design. HarperCollins e-books
14. Dolata M, Schwabe G (2016) Design Thinking in IS Research Projects. In: Brenner W, Uebernickel F (eds) Des. Think. Innov. Springer International Publishing, pp 67–83
15. Carleton T, Cockayne W, Tahvanainen A (2013) Playbook for strategic foresight and innovation.
16. Vetterli C, Hoffmann F, Brenner W, et al (2012) Designing innovation: Prototypes and team performance in design thinking. In: Proc. Conf. ISPIM, ISPIM, Barcelona, Spain.
17. Leifer LJ, Steinert M (2014) Dancing with Ambiguity: Causality Behavior, Design Thinking, and Triple-Loop-Learning. In: Manag. Fuzzy Front End Innov. Springer, pp 141–158
18. Mabogunje A, Leifer LJ, Levitt RE, Baudin C (1995) ME210-VDT: a managerial framework for measuring and improving design process performance. In: Front. Educ. Conf. 1995. IEEE.
19. Agogino A, Song S, Hey J (2007) Triangulation of indicators of successful student design teams. Int J Eng Educ 22:617.
20. Gero JS (1993) Modeling creativity and knowledge-based creative design. Erlbaum, Hillsdale, NJ, USA.
21. Kessel M, Kratzer J, Schultz C (2012) Psychological safety, knowledge sharing, and creative performance in healthcare teams. Creat Innov Manag 21:147–157.
22. Paulus P (2000) Groups, Teams, and Creativity: The Creative Potential of Idea-generating Groups. Appl Psychol 49:237–262. doi: 10.1111/1464-0597.00013
23. Paulus PB, Dzindolet M (2008) Social influence, creativity and innovation. Soc Influ 3:228–247.
24. Donati C, Vignoli M (2014) How tangible is your prototype? Designing the user and expert interaction. Int J Interact Des Manuf IJIDeM 1–8. doi: 10.1007/s12008-014-0232-5
25. Myers IB, McCaulley MH, Quenk NL, Hammer AL (1998) MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator. Cons. Psych. Press, Palo Alto.
26. Aronson ZH, Reilly RR, Lynn GS (2006) The impact of leader personality on new product development teamwork and performance. J Eng Technol Manag 23:221–247.
27. Barrick MR, Stewart GL, Neubert MJ, Mount MK (1998) Relating member ability and personality to work-team processes and team effectiveness. J Appl Psychol 83:377–391.
28. Bell ST (2007) Deep-level composition variables as predictors of team performance: A meta-analysis. J Appl Psychol 92:595–615.
29. 29. Kress GL, Schar M (2012) Applied Teamology: The Impact of Cognitive Style Diversity on Problem Reframing and Product Redesign Within Design Teams. In: Plattner H, Meinel C, Leifer L (eds) Des. Think. Res. Springer Berlin Heidelberg, pp 127–149
30. Kress GL, Schar M (2012) Teamology – The Art and Science of Design Team Formation. In: Plattner H, Meinel C, Leifer L (eds) Des. Think. Res. Springer, pp 189–209
31. Reilly RR, Lynn GS, Aronson ZH (2002) The role of personality in new product development team performance. J Eng Technol Manag 19:39–58.

32. Jung M, Chong J, Leifer L (2012) Group hedonic balance and pair programming performance: affective interaction dynamics as indicators of performance. In: Proc. SIGCHI Conf. Hum. Factors Comput. Syst. ACM, pp 829–838
33. Jung MF (2011) Engineering Team Performance and Emotion: Affective Interaction Dynamics as Indicators of Design Team Performance. Stanford University
34. Lande M, Sonalkar N, Jung M, et al (2012) Monitoring Design Thinking Through In-Situ Interventions. In: Plattner H, Meinel C, Leifer L (eds) Des. Think. Res. Springer, pp 211ff.
35. Sonalkar N, Jung M, Mabogunje A (2011) Emotion in Engineering Design Teams. In: Fukuda S (ed) Emot. Eng. Springer London, pp 311–326
36. Blackburn T, Swatman P (2007) Towards a framework for supporting professional teamwork: modelling human actions in small group meetings. Proc. Conf. ACIS.
37. Blackburn T, Swatman P, Vernik R (2007) Cognitive Dust: A Framework That Builds from CSCW Concepts to Provide Situated Support for Small Group Work. In: Shen W, Luo J, Lin Z, et al (eds) Comput. Support. Coop. Work Des. III. Springer, pp 1–12
38. Tan KL, Swatman P (2010) Modelling Creative Team Dynamics. In: Proc. Conf. HICSS. IEEE, pp 1–10
39. Kress G (2009) Multimodality: A Social Semiotic Approach to Contemporary Communication. Routledge, Abingdon; New York
40. Dong A (2009) The Language of Design. Springer London, London
41. Sonalkar N, Mabogunje A, Leifer L (2013) Developing a visual representation to characterize moment-to-moment concept generation in design teams. Int J Des Creat Innov 1:93–108.
42. Austin JL (1962) How to do things with words. Oxford university press
43. Vivas E (1955) Creation and Discovery Essays in Criticism and Aesthetics, Noonday Press
44. Lynch MD, Kays DJ (1967) Some Effects of Distribution of Writing Tasks and Creative Aptitude on Journalistic Performance. Journal Q 44:508–12.
45. Lynch MD, Swink E (1967) Some Effects of Priming, Incubation and Creative Aptitude on Journalism Performance. J Commun 17:372–382
46. Albert Á, Kormos J (2004) Creativity and Narrative Task Performance: An Exploratory Study. Lang Learn 54:277–310.
47. Baer J (1993) Creativity and Divergent Thinking: A Task-Specific Approach, 1 edition. Psychology Press, Hillsdale, N.J
48. Baer J (1996) The Effects of Task-Specific Divergent-Thinking Training. J Creat Behav 30:183–187.
49. Hollingshead AB, Mcgrath JE, O'Connor KM (1993) Group Task Performance and Communication Technology s. Small Group Res 24:307–333.
50. McGrath JE (1984) Groups: Interaction and performance.
51. Coughlan T, Johnson P (2009) Understanding Productive, Structural and Longitudinal Interactions in the Design of Tools for Creative Activities. In: Proc. Seventh ACM Conf. Creat. Cogn. ACM, New York, NY, USA, pp 155–164
52. Dolata M, Schwabe G (2014) Call for Action: Designing for Harmony in Creative Teams. In: Tremblay MC, VanderMeer DE, Rothenberger MA, et al (eds) Adv. Impact Des. Sci. Mov. Theory Pract., Springer, pp 273–288
53. Jehn KA, Mannix EA (2001) The Dynamic Nature of Conflict: A Longitudinal Study of Intragroup Conflict and Group Performance. Acad Manage J 44:238–251.
54. McCarthy PM, Jarvis S (2010) MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. Behav Res Methods 42:381–392.
55. Xanthos A (2014) Lingua::Diversity - measuring the diversity of text units. Avail. at CPAN